



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/587,990	06/06/2000	Chris A. Hamilton	024/1	8460

8791 7590 11/10/2003

BLAKELY SOKOLOFF TAYLOR & ZAFMAN
12400 WILSHIRE BOULEVARD, SEVENTH FLOOR
LOS ANGELES, CA 90025

EXAMINER

ENG, GEORGE

ART UNIT	PAPER NUMBER
----------	--------------

2643

23

DATE MAILED: 11/10/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary

Application No.

09/587,990

Applicant(s)

HAMILTON, CHRIS A.

Examiner

George Eng

Art Unit

2643

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 07 August 2003.
- 2a) ☐ This action is FINAL. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 5-29 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 5-29 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
- Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- 11) ☐ The proposed drawing correction filed on _____ is: a) ☐ approved b) ☐ disapproved by the Examiner.
- If approved, corrected drawings are required in reply to this Office action.
- 12) ☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

- 13) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- * See the attached detailed Office action for a list of the certified copies not received.
- 14) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).
- a) ☐ The translation of the foreign language provisional application has been received.
- 15) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☐ Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____
- 4) ☐ Interview Summary (PTO-413) Paper No(s). _____
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other:

DETAILED ACTION

Continued Examination Under 37 CFR 1.114

1. A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 8/7/2003 (paper no. 21) has been entered.

Response to Amendment

2. This office action is in response to the amendment filed 8/7/2003 (paper no. 22).

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 5-14, 19-22 and 26-29 are rejected under 35 U.S.C. 103(a) as being unpatentable over Ogata et al. (JP 06-062400 hereinafter Ogata) or Kamata et al. (US PAT. 5,953,050 hereinafter Kamata) in view of Zhou (US PAT. 5,550,580) and Taylor (EP 0254409 A1).

Regarding claim 5, Ogata discloses a conference control system comprising means (2, figure 1) for interconnecting a plurality of videoconference stations (1a-1f, figure 1). Kamata discloses a video conferencing system comprising means for interconnecting a plurality of videoconference stations (figure 1 and col. 1 lines 13-20). Ogata or Kamata differs from the claimed invention in not specifically teaching to determine whether a conferee is speaking by analyzing lip movements of said conferee with an audio signal from a conference station in which said conferee is located so as to produce human speech. However, Zhou teaches a lip motion subroutine for detecting the location and movement of the lips of a person present in video scene with an audio signal in order to accurately indicate human speech (abstract, col. 2 lines 1-47, col. 17 line 36 through col. 18 line 59 and col. 22 lines 5-19). Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to modify either Ogata or Kamata in having the algorithm for determining whether the conferee is speaking by analyzing lip movements of said conferee with an audio signal from a conference station in which said conferee is located so as to produce human speech, as per teaching of Zhou, because it improves perceptual quality so that the audio signal will be encoded with greater accuracy than the video signal when the audio signal is correlated with lip movements. Furthermore, the combination of Ogata or Kamata and Zhuo differs from the claimed invention in not specifically teaching to analyze a consistency between visual lip movements with the audio signal from the conference station. However, Taylor teaches a technique of recognizing speech signal by analyzing a consistency between information derived from a camera, i.e., lip movements, with audio signal derived from a microphone in order to reliably identify speech sound in noisy environment (col. 1 line 52 through col. 2 line 7 and col. 2 line 39 through col. 5

Art Unit: 2643

line 34). Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the combination of Ogata or Kamata and Zhuo in analyzing the consistency between visual lip movements with the audio signal from the conference station, as per teaching of Taylor, in order to identify speech sound in noisy environment.

Regarding claims 6-8, Ogata disclose to identify the presence or absence of speech of each participant according to the voice level of the conference participant (abstract). Thus, a voice activity detector is obviously located at each conference stations or implemented at the conference bridge. In addition, Kamata also teaches that means for altering is responsive to a voice activity detector (76) located at each conference stations or implemented at the conference bridge (figure 12 and col. 12 lines 4-13).

Regarding claim 9, Zhou teaches image analysis and recognition software (col. 13 line 22 through col. 15 line 45).

Regarding claim 10, Ogata teaches to display a red rectangular marker in a window display frame to indicate who is a speaker (abstract). In addition, Kamata also discloses means for emphasizing an image of a remote speaker to be speaking (fig. 2B and col. 1 lines 36-41).

Regarding claim 11, Ogata discloses a videoconference station (1a, figure 1) obviously comprising a transmitter to transmit a combined audio and video signal to a videoconference bridge (abstract). Kamata discloses a videoconference station (1, figure 1) obviously comprising a transmitter to transmit a combined audio and video signal to a videoconference bridge (figure 1 and col. 1 lines 36-41). Ogata or Kamata differs from the claimed invention in not specifically teaching that an algorithm for determining whether a conferee is speaking by analyzing lip movements of said conferee with an audio signal from a conference station in which said

Art Unit: 2643

conferee is located so as to produce human speech. However, Zhou teaches a lip motion subroutine for detecting the location and movement of the lips of a person present in video scene with an audio signal in order to accurately indicate human speech (abstract, col. 2 lines 1-47, col. 17 line 36 through col. 18 line 59 and col.22 lines 5-19). Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to modify either Ogata or Kamata in having the algorithm for determining whether the conferee is speaking by analyzing lip movements of said conferee with an audio signal from a conference station in which said conferee is located so as to produce human speech, as per teaching of Zhou, because it improves perceptual quality so that the audio signal will be encoded with greater accuracy than the video signal when the audio signal is correlated with lip movements. Furthermore, the combination of Ogata or Kamata and Zhuo differs from the claimed invention in not specifically teaching to analyze a consistency between visual lip movements with the audio signal from the conference station. However, Taylor teaches a technique of recognizing speech signal by analyzing a consistency between information derived from a camera, i.e., lip movements, with audio signal derived from a microphone in order to reliably identify speech sound in noisy environment (col. 1 line 52 through col. 2 line 7 and col. 2 line 39 through col. 5 line 34). Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the combination of Ogata or Kamata and Zhuo in analyzing the consistency between visual lip movements with the audio signal from the conference station, as per teaching of Taylor, in order to identify speech sound in noisy environment.

Regarding claim 12, Ogata disclose to identify the presence or absence of speech of each participant according to the voice level of the conference participant (abstract). Thus, a voice

Art Unit: 2643

activity detector is obviously located at videoconference station. In addition, Kamata also teaches that means for altering is responsive to a voice activity detector (76) located at conference station (figure 12 and col. 12 lines 4-13).

Regarding claim 13, the limitations of the claim are rejected as the same reasons set forth in claim 9.

Regarding claims 14, Ogata discloses a conference control system comprising means (2, figure 1) for interconnecting a plurality of videoconference stations (1a-1f, figure 1) and means for visually altering an image of at least one of a plurality of remotely located conferees who is a speaker at a particular time (abstract). Kamata discloses a video conferencing system comprising means for interconnecting a plurality of video conference stations (figure 1 and col. 1 lines 13-20) and means for visually altering an image of at least one of a plurality of remotely located conferees when said of at least one of said plurality of remotely located conferees is speaking (figure 2B and col. 1 lines 36-41). Ogata or Kamata differs from the claimed invention in not specifically teaching to determine whether a conferee is speaking by analyzing lip movements of said conferee with an audio signal from a conference station in which said conferee is located so as to produce human speech. However, Zhou teaches a lip motion subroutine for detecting the location and movement of the lips of a person present in video scene with an audio signal in order to accurately indicate human speech (abstract, col. 2 lines 1-47, col. 17 line 36 through col. 18 line 59 and col. 22 lines 5-19). Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to modify either Ogata or Kamata in having the algorithm for determining whether the conferee is speaking by analyzing lip movements of said conferee with an audio signal from a conference station in which said conferee is located so as to

Art Unit: 2643

produce human speech, as per teaching of Zhou, because it improves perceptual quality so that the audio signal will be encoded with greater accuracy than the video signal when the audio signal is correlated with lip movements. Furthermore, the combination of Ogata or Kamata and Zhuo differs from the claimed invention in not specifically teaching to analyze a consistency between visual lip movements with the audio signal from the conference station. However, Taylor teaches a technique of recognizing speech signal by analyzing a consistency between information derived from a camera, i.e., lip movements, with audio signal derived from a microphone in order to reliably identify speech sound in noisy environment (col. 1 line 52 through col. 2 line 7 and col. 2 line 39 through col. 5 line 34). Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the combination of Ogata or Kamata and Zhuo in analyzing the consistency between visual lip movements with the audio signal from the conference station, as per teaching of Taylor, in order to identify speech sound in noisy environment.

Regarding claim 19, Ogata discloses a system for identifying which conferee in a videoconference is speaking comprising the step of providing an indication to the first conferee and the second conferee of which detected audio signal is louder (abstract), as well as Kamata (fig. 2B and col. 1 lines 36-41). Ogata or Kamata differs from the claimed invention in not specifically teaching that an algorithm for determining whether a conferee is speaking by analyzing lip movements of said conferee with an audio signal from a conference station in which said conferee is located so as to produce human speech. However, Zhou teaches a lip motion subroutine for detecting the location and movement of the lips of a person present in video scene with an audio signal in order to accurately indicate human speech (abstract, col. 2

Art Unit: 2643

lines 1-47, col. 17 line 36 through col. 18 line 59 and col.22 lines 5-19). Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to modify either Ogata or Kamata in having the algorithm for determining whether the conferee is speaking by analyzing lip movements of said conferee with an audio signal from a conference station in which said conferee is located so as to produce human speech, as per teaching of Zhou, because it improves perceptual quality so that the audio signal will be encoded with greater accuracy than the video signal when the audio signal is correlated with lip movements. Furthermore, the combination of Ogata or Kamata and Zhuo differs from the claimed invention in not specifically teaching to analyze a consistency between visual lip movements with the audio signal from the conference station. However, Taylor teaches a technique of recognizing speech signal by analyzing a consistency between information derived from a camera, i.e., lip movements, with audio signal derived from a microphone in order to reliably identify speech sound in noisy environment (col. 1 line 52 through col. 2 line 7 and col. 2 line 39 through col. 5 line 34). Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the combination of Ogata or Kamata and Zhuo in analyzing the consistency between visual lip movements with the audio signal from the conference station, as per teaching of Taylor, in order to identify speech sound in noisy environment.

Regarding claim 20, Ogata discloses means for visually altering an image of at least one of a plurality of remotely located conferees who is a speaker at a particular time (abstract), as well as Kamata (figure 2B and col. 1 lines 36-41).

Regarding claims 21-22, the limitations of the claims are rejected as the same reasons set forth in claim 10.

Art Unit: 2643

Regarding claim 26, the limitations of the claim are rejected as the same reasons set forth in claim 5.

Regarding claim 27, the limitations of the claim are rejected as the same reasons set forth in claim 9.

Regarding claim 28, Ogata discloses a display unit for providing visual representation of conferees participating in a videoconference (figure 2), as well as Kamata (figures 2A-2B).

Regarding claim 29, the limitations of the claim are rejected as the same reasons set forth in claim 20.

5. Claims 15-16 and 23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Zhou (US PAT. 5,550,580) in view of Taylor (EP 0254409 A1).

Regarding claim 15, Zhou teaches a method for determining whether a conferee in a videoconference is speaking, comprising analyzing visual lip movements of said conferee with an audio signal from a conference station in which said conferee is located such that the combination of lip movements and audio signal indicates human speech (abstract, col. 2 lines 1-47, col. 17 line 36 through col. 18 line 59 and col. 22 lines 5-19). Zhou differs from the claimed invention in not specifically teaching to analyze a consistency between visual lip movements with the audio signal from the conference station. However, Taylor teaches a technique of recognizing speech signal by analyzing a consistency between information derived from a camera, i.e., lip movements, with audio signal derived from a microphone in order to reliably identify speech sound in noisy environment (col. 1 line 52 through col. 2 line 7 and col. 2 line 39 through col. 5 line 34). Therefore, it would have been obvious to a person of ordinary skill in the

Art Unit: 2643

art at the time the invention was made to modify Zhuo in analyzing the consistency between visual lip movements with the audio signal from the conference station, as per teaching of Taylor, in order to identify speech sound in noisy environment.

Regarding claims 16 and 23, the limitations of the claim are rejected as the same reasons set forth in claim 15.

6. Claims 17-18 and 24-25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Zhou (US PAT. 5,550,580) in view of Taylor (EP 0254409 A1) as applied in claims above, and further in view of Ogata et al. (JP 06-062400 hereinafter Ogata) or Kamata et al. (US PAT. 5,953,050 hereinafter Kamata).

Regarding claims 17-18, the combination of Zhou and Taylor differs from the claimed invention in not specifically teaching to alter an image of the conferee that is display to other conferees if the conferees is determined to be speaking and to provide textual information or highlighting a border around the image to identify the conferee to other conferees if the conferee is determined to be speaking. However, means for visually altering an image of at least one of a plurality of remotely located conferees who is a speaker at a particular time (abstract) and means for displaying a red rectangular marker in a window display frame to indicate who is a speaker (abstract), as well as Kamata (figure 2B and col. 1 lines 36-41), in order to make user friendly by providing visual notification to the conferees when a speaker is determined. Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the combination of Zhou and Taylor in altering the image of the conferee that is display to other conferees if the conferees is determined to be speaking and providing textual

Art Unit: 2643

information or highlighting a border around the image to identify the conferee to other conferees if the conferee is determined to be speaking, as per teaching of Ogata and Kamata, in order to make user friendly.

Response to Arguments

7. Applicant's arguments with respect to claims 5-29 have been considered but are moot in view of the new ground(s) of rejection.

Conclusion

8. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Duttweiler et al. (US PAT. 5,818,514) discloses a video conferencing system and method for providing enhanced interactive communication (abstract). Cooper (US PAT. 5,572,261) discloses a method for detecting or measuring relative audio and video timing in audio-visual communications system (abstract).

9. Any response to this action should be mailed to:

Commissioner of Patents and Trademarks

Washington D.C. 20231

Or faxed to:

(703) 872-9306 (for Technology Center 2600 only)

Hand delivered responses should be brought to Crystal Park II, 2121 Crystal Drive, Arlington, V.A., Sixth Floor (Receptionist).

Art Unit: 2643

10. Any inquiry concerning this communication or earlier communications from the examiner should be directed to George Eng whose telephone number is 703-308-9555. The examiner can normally be reached on Tuesday to Friday from 7:30 AM to 6:00 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Curtis A. Kuntz, can be reached on (703) 305-4870. The fax phone number for the organization where this application or proceeding is assigned is 703-308-6306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703) 306-0377.

A handwritten signature in cursive script that reads "George Eng".

George Eng

Primary Examiner

Art Unit 2643